



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

| | | |
|---|-----------|---|
| (51) International Patent Classification ⁷ : C07K 14/435, C12N 1/00, 1/19, 5/10, 15/12, 15/63 | A1 | (11) International Publication Number: WO 00/34321 (43) International Publication Date: 15 June 2000 (15.06.00) |
| (21) International Application Number: PCT/US99/29402 (22) International Filing Date: 10 December 1999 (10.12.99) (30) Priority Data: 09/210,330 11 December 1998 (11.12.98) US 09/418,917 15 October 1999 (15.10.99) US (71) Applicant: CLONTECH LABORATORIES, INC. [US/US]; 1020 East Meadow Drive, Palo Alto, CA 94303 (US). (72) Inventors: LUKYANOY, Sergey Anatolievich; ul. Golubinskaya 13/1-161, Moscow (RU). FRADKOV, Arcady Fedorovich; ul. Dnepropetrovskaya, 35/2-14, Moscow, 113570 (RU). LABAS, Yulii Aleksandrovich; ul. Generala Tyuleneva, 35/416, Moscow, 117465 (RU). MATZ, Mikhail Vladimirovich; ul. Teplii stan, 7/2-28, Moscow, 117465 (RU). ANGRES, Brigitte; 10534 Sterling Boulevard, Cupertino, CA 95014 (US). GREEN, Gisele; 4106 Thain Way, Palo Alto, CA 94306 (US). (74) Agent: ADLER, Benjamin, A.; McGregor & Adler, 8011 Candle Ln., Houston, TX 77071 (US). | | (81) Designated States: JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i> |
| (54) Title: FLUORESCENT PROTEINS FROM NON-BIOLUMINESCENT SPECIES OF CLASS ANTHOZOA, GENES ENCODING SUCH PROTEINS AND USES THEREOF | | |
| (57) Abstract The present invention is directed to novel fluorescent proteins from non-bioluminescent organisms from the Class Anthozoa. Also disclosed are cDNAs encoding the fluorescent proteins. | | |

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|----|--------------------------|----|--|----|--|----|--------------------------|
| AL | Albania | ES | Spain | LS | Lesotho | SI | Slovenia |
| AM | Armenia | FI | Finland | LT | Lithuania | SK | Slovakia |
| AT | Austria | FR | France | LU | Luxembourg | SN | Senegal |
| AU | Australia | GA | Gabon | LV | Latvia | SZ | Swaziland |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BB | Barbados | GH | Ghana | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan |
| BF | Burkina Faso | GR | Greece | | | TR | Turkey |
| BG | Bulgaria | HU | Hungary | ML | Mali | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MN | Mongolia | UA | Ukraine |
| BR | Brazil | IL | Israel | MR | Mauritania | UG | Uganda |
| BY | Belarus | IS | Iceland | MW | Malawi | US | United States of America |
| CA | Canada | IT | Italy | MX | Mexico | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NE | Niger | VN | Viet Nam |
| CG | Congo | KE | Kenya | NL | Netherlands | YU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NO | Norway | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | Democratic People's Republic of Korea | NZ | New Zealand | | |
| CM | Cameroon | | | PL | Poland | | |
| CN | China | KR | Republic of Korea | PT | Portugal | | |
| CU | Cuba | KZ | Kazakhstan | RO | Romania | | |
| CZ | Czech Republic | LC | Saint Lucia | RU | Russian Federation | | |
| DE | Germany | LI | Liechtenstein | SD | Sudan | | |
| DK | Denmark | LK | Sri Lanka | SE | Sweden | | |
| EE | Estonia | LR | Liberia | SG | Singapore | | |

**FLUORESCENT PROTEINS FROM NON-BIOLUMINESCENT SPECIES
OF CLASS ANTHOZOA, GENES ENCODING SUCH PROTEINS AND
USES THEREOF**

5

BACKGROUND OF THE INVENTION

10

Cross-reference to Related Application

This is a divisional application of U.S.S.N. 09/210,330 filed on December 11, 1998.

15 Field of the Invention

This invention relates to the field of molecular biology. More specifically, this invention relates to novel fluorescent proteins, cDNAs encoding the proteins and uses thereof.

20 Description of the Related Art

Fluorescence labeling is a particularly useful tool for marking a protein, cell, or organism of interest. Traditionally, a protein of interest is purified, then covalently conjugated to a fluorophore derivative. For *in vivo* studies, the protein-dye complex is
25 then inserted into cells of interest using micropipetting or a method of reversible permeabilization. The dye attachment and insertion steps, however, make the process laborious and difficult to control. An alternative method of labeling proteins of interest is to concatenate or fuse the gene expressing the protein of interest to a gene expressing a

marker, then express the fusion product. Typical markers for this method of protein labeling include β -galactosidase, firefly luciferase and bacterial luciferase. These markers, however, require exogenous substrates or cofactors and are therefore of limited use for *in vivo* studies.

A marker that does not require an exogenous cofactor or substrate is the green fluorescent protein (GFP) of the jellyfish *Aequorea victoria*, a protein with an excitation maximum at 395 nm, a second excitation peak at 475 nm and an emission maximum at 510 nm. GFP is a 238-amino acid protein, with amino acids 65-67 involved in the formation of the chromophore.

Uses of GFP for the study of gene expression and protein localization are discussed in detail by Chalfie et al. in *Science* 263 (1994), 802-805, and Heim et al. in *Proc. Nat. Acad. Sci.* 91 (1994), 12501-12504. Additionally, Rizzuto et al. in *Curr. Biology* 5 (1995), 635-642, discuss the use of wild-type GFP as a tool for visualizing subcellular organelles in cells, while Kaether and Gerdes in *Febs Letters* 369 (1995), 267-271, report the visualization of protein transport along the secretory pathway using wild-type GFP. The expression of GFP in plant cells is discussed by Hu and Cheng in *Febs Letters* 369 (1995), 331-334, while GFP expression in *Drosophila* embryos is described by Davis et al. in *Dev. Biology* 170 (1995), 726-729.

Crystallographic structures of wild-type GFP and the mutant GFP S65T reveal that the GFP tertiary structure resembles a barrel (Ormö et al., *Science* 273 (1996), 1392-1395; Yang, et al., *Nature Biotechnol* 14 (1996), 1246-1251). The barrel consists of beta sheets in a compact structure, where, in the center, an alpha helix containing the chromophore is shielded by the barrel. The compact structure makes GFP very stable under diverse and/or harsh conditions such as

protease treatment, making GFP an extremely useful reporter in general. However, the stability of GFP makes it sub-optimal for determining short-term or repetitive events.

A great deal of research is being performed to improve the properties of GFP and to produce GFP reagents useful and optimized for a variety of research purposes. New versions of GFP have been developed, such as a "humanized" GFP DNA, the protein product of which has increased synthesis in mammalian cells (Haas, et al., *Current Biology* 6 (1996), 315-324; Yang, et al., *Nucleic Acids Research* 24 (1996), 4592-4593). One such humanized protein is "enhanced green fluorescent protein" (EGFP). Other mutations to GFP have resulted in blue-, cyan- and yellow-green light emitting versions. Despite the great utility of GFP, however, other fluorescent proteins with properties similar to or different from GFP would be useful in the art. Novel fluorescent proteins result in possible new colors, or produce pH-dependent fluorescence. Other benefits of novel fluorescent proteins include fluorescence resonance energy transfer (FRET) possibilities based on new spectra and better suitability for larger excitation.

The prior art is deficient in novel fluorescent proteins wherein the DNA coding sequences are known. The present invention fulfills this long-standing need in the art.

SUMMARY OF THE INVENTION

25

The present invention is directed to DNA sequences encoding fluorescent proteins selected from the group consisting of: (a) an isolated DNA from an organism from the Class Anthozoa which encodes a fluorescent protein; (b) an isolated DNA which hybridizes to

the isolated DNA of (a) and which encodes a fluorescent protein; and
(c) an isolated DNA differing from the isolated DNAs of (a) and (b) in
codon sequence due to the degeneracy of the genetic code and that
encodes a fluorescent protein. Preferably, the DNA is isolated from a
5 non-bioluminescent organism from Class Anthozoa. More preferably,
the DNA has the sequence shown in SEQ ID No. 55 and the fluorescent
protein has the amino acid sequence shown in SEQ ID No. 56.

In another embodiment of the present invention, there is
provided a vector capable of expressing the DNA of the present
10 invention in a recombinant cell comprising the DNA and regulatory
elements necessary for expression of the DNA in the cell. Preferably,
the DNA encodes a fluorescent protein having the amino acid sequence
shown in SEQ ID No. 56.

In still another embodiment of the present invention, there
15 is provided a host cell transfected with a vector of the present
invention, such that the host cell expresses a fluorescent protein.
Preferably, the cell is selected from the group consisting of bacterial
cells, mammalian cells, plant cells and insect cells. A representative
example of bacterial cell is an *E. coli* cell.

20 The present invention is also directed to an isolated and
purified fluorescent protein coded for by DNA selected from the group
consisting of: (a) isolated DNA from an organism from Class Anthozoa
which encodes a fluorescent protein; (b) isolated DNA which hybridizes
to the isolated DNA of (a) and which encodes a fluorescent protein;
25 and (c) isolated DNA differing from the isolated DNAs of (a) and (b) in
codon sequence due to the degeneracy of the genetic code, and which
encodes a fluorescent protein. Preferably, the protein has the amino
acid sequence shown in SEQ ID No. 56.

The present invention is also directed to a DNA sequence encoding a fluorescent protein selected from the group consisting of: (a) an isolated DNA which encodes a fluorescent protein, wherein the DNA is from an organism from Class Anthozoa and wherein the organism does not exhibit bioluminescence; (b) an isolated DNA which hybridizes to isolated DNA of (a) and which encodes a fluorescent protein; and (c) an isolated DNA differing from the isolated DNAs of (a) and (b) in codon sequence due to degeneracy of the genetic code and which encodes a fluorescent protein. Preferably, the organism is from Sub-class-Alcyonaria, Order Stolonifera. More preferably, the organism is from Family Clavulariidae, Genus Clavularia. Most particularly, the present invention is drawn to a novel fluorescent protein from *Clavularia* sp., cFP484.

The present invention is further directed to an amino acid sequence which can be used as a basis for designing an oligonucleotide probe for identification of a DNA encoding a fluorescent protein by means of hybridization, wherein the amino acid sequence is selected from the group consisting of SEQ ID Nos. 3, 5, 8, 11, 12, 14. Preferably, such an oligonucleotide has a nucleotide sequence selected from the group consisting of SEQ ID Nos. 4, 6, 7, 9, 10, 13, 15, 16.

Other and further aspects, features, and advantages of the present invention will be apparent from the following description of the presently preferred embodiments of the invention given for the purpose of disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the modified strategy of 3'-RACE used to isolate the target fragments. Sequences of the oligonucleotides used are shown in Table 2. Dp1 and Dp2 are the degenerate primers used in the first and second PCR, respectively (see Tables 3 and 4 for the sequences of degenerate primers). In the case of *Clavularia sp.*, the first degenerate primer used was NGH (SEQ ID No. 4), and the second degenerate primer used was GEG(a) (SEQ ID No. 6).

Figure 2 shows the excitation and emission spectrum of the novel fluorescent protein from *Clavularia sp.*, cFP484.

Figure 3 shows transient expression of cFP484 lacking the N-terminal 19 amino acids ($\Delta 19$ cFP484) in mammalian cells. Figure 3A shows the image emitted by $\Delta 19$ cFP484 using the filter set XF 114 (Omega Optical). Figure 3B shows phase contrast image taken from the same field of view as that in Figure 3A and overlaid with the image in Figure 3A.

DETAILED DESCRIPTION OF THE INVENTION

As used herein, the term "GFP" refers to the basic green fluorescent protein from *Aequorea victoria*, including prior art versions of GFP engineered to provide greater fluorescence or fluoresce in different colors. The sequence of *Aequorea victoria* GFP (SEQ ID No. 54) has been disclosed in Prasher et al., *Gene* 111 (1992), 229-33.

As used herein, the term "EGFP" refers to mutant variant of GFP having two amino acid substitutions: F64L and S65T (Heim et al., *Nature* 373 (1995), 663-664). The term "humanized" refers to changes made to the GFP nucleic acid sequence to optimize the codons for

expression of the protein in human cells (Yang et al., *Nucleic Acids Research* 24 (1996), 4592-4593).

In accordance with the present invention there may be employed conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. See, e.g., Maniatis, Fritsch & Sambrook, "Molecular Cloning: A Laboratory Manual (1982); "DNA Cloning: A Practical Approach," Volumes I and II (D.N. Glover ed. 1985); "Oligonucleotide Synthesis" (M.J. Gait ed. 1984); "Nucleic Acid Hybridization" (B.D. Hames & S.J. Higgins eds. (1985)); "Transcription and Translation" (B.D. Hames & S.J. Higgins eds. (1984)); "Animal Cell Culture" (R.I. Freshney, ed. (1986)); "Immobilized Cells and Enzymes" (IRL Press, (1986)); B. Perbal, "A Practical Guide To Molecular Cloning" (1984).

A "vector" is a replicon, such as plasmid, phage or cosmid, to which another DNA segment may be attached so as to bring about the replication of the attached segment.

A "DNA molecule" refers to the polymeric form of deoxyribonucleotides (adenine, guanine, thymine, or cytosine) in either single stranded form or a double-stranded helix. This term refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms. Thus, this term includes double-stranded DNA found, inter alia, in linear DNA molecules (e.g., restriction fragments), viruses, plasmids, and chromosomes.

A DNA "coding sequence" is a DNA sequence which is transcribed and translated into a polypeptide in vivo when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3'

(carboxyl) terminus. A coding sequence can include, but is not limited to, prokaryotic sequences, cDNA from eukaryotic mRNA, genomic DNA sequences from eukaryotic (e.g., mammalian) DNA, and synthetic DNA sequences. A polyadenylation signal and transcription termination
5 sequence may be located 3' to the coding sequence.

As used herein, the term "hybridization" refers to the process of association of two nucleic acid strands to form an antiparallel duplex stabilized by means of hydrogen bonding between residues of the opposite nucleic acid strands.

10 — The term "oligonucleotide" refers to a short (under 100 bases in length) nucleic acid molecule.

"DNA regulatory sequences", as used herein, are transcriptional and translational control sequences, such as promoters, enhancers, polyadenylation signals, terminators, and the like, that
15 provide for and/or regulate expression of a coding sequence in a host cell.

A "promoter sequence" is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. For purposes of defining
20 the present invention, the promoter sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be found a
25 transcription initiation site, as well as protein binding domains responsible for the binding of RNA polymerase. Eukaryotic promoters will often, but not always, contain "TATA" boxes and "CAT" boxes. Various promoters, including inducible promoters, may be used to drive the various vectors of the present invention.

As used herein, the terms "restriction endonucleases" and "restriction enzymes" refer to bacterial enzymes, each of which cut double-stranded DNA at or near a specific nucleotide sequence.

A cell has been "transformed" or "transfected" by exogenous or heterologous DNA when such DNA has been introduced inside the cell. The transforming DNA may or may not be integrated (covalently linked) into the genome of the cell. In prokaryotes, yeast, and mammalian cells for example, the transforming DNA may be maintained on an episomal element such as a plasmid. With respect to eukaryotic cells, a stably transformed cell is one in which the transforming DNA has become integrated into a chromosome so that it is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter cells containing the transforming DNA. A "clone" is a population of cells derived from a single cell or common ancestor by mitosis. A "cell line" is a clone of a primary cell that is capable of stable growth *in vitro* for many generations.

A "heterologous" region of the DNA construct is an identifiable segment of DNA within a larger DNA molecule that is not found in association with the larger molecule in nature. Thus, when the heterologous region encodes a mammalian gene, the gene will usually be flanked by DNA that does not flank the mammalian genomic DNA in the genome of the source organism. In another example, heterologous DNA includes coding sequence in a construct where portions of genes from two different sources have been brought together so as to produce a fusion protein product. Allelic variations or naturally-occurring mutational events do not give rise to a heterologous region of DNA as defined herein.

As used herein, the term "reporter gene" refers to a coding sequence attached to heterologous promoter or enhancer elements and whose product may be assayed easily and quantifiably when the construct is introduced into tissues or cells.

5 The amino acids described herein are preferred to be in the "L" isomeric form. The amino acid sequences are given in one-letter code (A: alanine; C: cysteine; D: aspartic acid; E: glutamic acid; F: phenylalanine; G: glycine; H: histidine; I: isoleucine; K: lysine; L: leucine; M: methionine; N: asparagine; P: proline; Q: glutamine; R: arginine; S:
10 serine; T: threonine; V: valine; W: tryptophan; Y: tyrosine; X: any residue). NH₂ refers to the free amino group present at the amino terminus of a polypeptide. COOH refers to the free carboxy group present at the carboxy terminus of a polypeptide. In keeping with standard polypeptide nomenclature, *J Biol. Chem.*, 243 (1969), 3552-
15 59 is used.

 The present invention is directed to an isolated DNA selected from the group consisting of: (a) isolated DNA from an organism from the Class Anthozoa which encodes a fluorescent protein; (b) isolated DNA which hybridizes to isolated DNA of (a) and
20 which encodes a fluorescent protein; and (c) isolated DNA differing from the isolated DNAs of (a) and (b) in codon sequence due to the degeneracy of the genetic code, and which encodes a fluorescent protein. Preferably, the DNA has the sequence shown in SEQ ID No. 55 and the fluorescent protein has the amino acid sequence shown in SEQ
25 ID No. 56.

 In another embodiment of the present invention, there is provided a vector capable of expressing the DNA of the present invention in a recombinant cell comprising the DNA and regulatory elements necessary for expression of the DNA in the cell. Specifically,

the DNA encodes a fluorescent protein having the amino acid sequence shown in SEQ ID No. 56.

In still another embodiment of the present invention, there is provided a host cell transfected with the vector of the present invention, which expresses a fluorescent protein of the present invention. Preferably, the cell is selected from the group consisting of bacterial cells, mammalian cells, plant cells and insect cells. A representative example of bacterial cell is an *E. coli* cell.

The present invention is also directed to a DNA sequence encoding a fluorescent protein selected from the group consisting of: (a) an isolated DNA which encodes a fluorescent protein, wherein the DNA is from an organism from Class Anthozoa and wherein the organism does not exhibit bioluminescence; (b) an isolated DNA which hybridizes to isolated DNA of (a) and which encodes a fluorescent protein; and (c) an isolated DNA differing from the isolated DNAs of (a) and (b) in codon sequence due to degeneracy of the genetic code and which encodes a fluorescent protein. Preferably, the organism is from Sub-class Alcyonaria, Order Stolonifera. More preferably, the organism is from Family Clavulariidae, Genus Clavularia.

The present invention is also directed to an isolated and purified fluorescent protein coded for by DNA selected from the group consisting of: (a) an isolated protein encoded by a DNA which encodes a fluorescent protein wherein the DNA is from an organism from Class Anthozoa and wherein the organism does not exhibit bioluminescence; (b) an isolated protein encoded by a DNA which hybridizes to isolated DNA of (a); and (c) an isolated protein encoded by a DNA differing from the isolated DNAs of (a) and (b) in codon sequence due to degeneracy of the genetic code. Preferably, the isolated and purified fluorescent protein is cFP484.

The present invention is further directed to an amino acid sequence which can be used as a basis for designing an oligonucleotide probe for identification of a DNA encoding a fluorescent protein by means of hybridization, wherein the amino acid sequence is selected from the group consisting of SEQ ID Nos. 3, 5, 8, 11, 12, 14. Preferably, such an oligonucleotide has a nucleotide sequence selected from the group consisting of SEQ ID Nos. 4, 6, 7, 9, 10, 13, 15, 16 and is used as a primer in polymerase chain reaction. Alternatively, it can be used as a probe for hybridization screening of the cloned genomic or cDNA library.

The following examples are given for the purpose of illustrating various embodiments of the invention and are not meant to limit the present invention in any fashion.

15

EXAMPLE 1

Biological Material

Novel fluorescent proteins were identified from several genera of Anthozoa which do not exhibit any bioluminescence but have fluorescent color as observed under usual white light or ultraviolet light. Six species were chosen (see Table 1).

TABLE 1Anthozoa Species Used in This Study

| Species | Area of Origination | Fluorescent Color |
|----------------------------|---------------------|---|
| Anemonia majano | Western Pacific | bright green tentacle tips |
| Clavularia sp. | Western Pacific | bright green tentacles and oral disk |
| Zoanthus sp. | Western Pacific | green-yellow tentacles and oral disk |
| Discosoma sp. "red" | Western Pacific | orange-red spots oral disk |
| Discosoma striata | Western Pacific | blue-green stripes on oral disk |
| Discosoma sp. "magenta" | Western Pacific | faintly purple oral disk |
| Discosoma sp. "green" | Western Pacific | green spots on oral disk |
| Anemonia sulcata | Mediterranean | purple tentacle tips |

EXAMPLE 2cDNA Preparation

Total RNA was isolated from the species of interest according to the protocol of Chomczynski and Sacchi (Chomczynski P., et al., *Anal. Biochem.* 162 (1987), 156-159). First-strand cDNA was synthesized starting with 1-3 µg of total RNA using SMART PCR cDNA synthesis kit (CLONTECH) according to the provided protocol with the only alteration being that the "cDNA synthesis primer" provided in the kit was replaced by the primer TN3 (5'- CGCAGTCGACCG(T)₁₃, SEQ ID No. 1) (Table 2). Amplified cDNA samples were then prepared as described in the protocol provided except the two primers used for PCR were the TS primer (5'-AAGCAGTGGTATCAACGCAGAGT, SEQ ID No. 2) (Table 2) and the TN3 primer (Table 2), both in 0.1 µM concentration. Twenty to twenty-five PCR cycles were performed to amplify a cDNA sample. The amplified cDNA was diluted 20-fold in water and 1 µl of this dilution was used in subsequent procedures.

TABLE 2Oligos Used in cDNA Synthesis and RACE

5 TN3: 5'-CGCAGTCGACCG(T)₁₃
(SEQ ID No. 1)

T7-TN3: 5'-GTAATACGACTCACTATAGGGCCGCAGTCGACCG(T)₁₃
(SEQ ID No. 17)

10

TS-primer: 5'-AAGCAGTGGTATCAACGCAGAGT
(SEQ ID No. 2)

T7-TS:
15 5'-GTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT
(SEQ ID No. 18)

T7: 5'-GTAATACGACTCACTATAGGGC
(SEQ ID No. 19)

20

TS-oligo 5'-AAGCAGTGGTATCAACGCAGAGTACGCrGrGrG
(SEQ ID No. 53)

25

EXAMPLE 3Oligo Design

To isolate fragments of novel fluorescent protein cDNAs,
5 PCR using degenerate primers was performed. Degenerate primers
were designed to match the sequence of the mRNAs in regions that
were predicted to be the most invariant in the family of fluorescent
proteins. Four such stretches were chosen (Table 3) and variants of
degenerate primers were designed. All such primers were directed to
10 the 3'-end of mRNA. All oligos were gel-purified before use. Table 2
shows the oligos used in cDNA synthesis and RACE.

TABLE 3

Key Amino Acid Stretches and Corresponding Degenerate Primers Used for Isolation of Fluorescent Proteins

5

| Stretch Position according to A. victoria GFP (7) | Amino Acid Sequence of the Key Stretch | Degenerated Primer Name and Sequence |
|---|--|---|
| 20-25 — | GXVNGH (SEQ ID No. 3) | NGH: 5'- GA(C,T) GGC TGC GT(A,T,G,C) AA(T,C) GG(A,T,G) CA (SEQ ID No. 4) |
| 31-35 | GEGEG (SEQ ID No. 5) GEGNG (SEQ ID No. 8) | GEGa: 5'- GTT ACA GGT GA(A,G) GG(A,C) GA(A,G) GG (SEQ ID No. 6) GEGb: 5'- GTT ACA GGT GA(A,G) GG(T,G) GA(A,G) GG (SEQ ID No. 7) GNGa: 5'- GTT ACA GGT GA(A,G) GG(A,C) AA(C,T) GG (SEQ ID No. 9) GNGb: 5'- GTT ACA GGT GA(A,G) GG(T,G) AA(C,T) GG (SEQ ID No. 10) |
| 127-131 | GMNFP (SEQ ID No. 11) GVNFP (SEQ ID No. 12) | NFP: 5' TTC CA(C,T) GGT (G,A)TG AA(C,T) TT(C,T) CC (SEQ ID NO. 13) |
| 134-137 | GPVM (SEQ ID No. 14) | PVMa: 5' CCT GCC (G,A)A(C,T) GGT CC(A,T,G,C) GT(A,C) ATG (SEQ ID NO. 15) PVMb: 5' CCT GCC (G,A)A(C,T) GGT CC(A,T,G,C) GT(G,T) ATG (SEQ ID NO. 16) |

EXAMPLE 4Isolation of 3'-cDNA Fragments of nFPs

The modified strategy of 3'-RACE was used to isolate the target fragments (see Figure 1). The RACE strategy involved two consecutive PCR steps. The first PCR step involved a first degenerate primer (Table 4) and the T7-TN3 primer (SEQ ID No. 17) which has a 3' portion identical to the TN3 primer used for cDNA synthesis (for sequence of T7-TN3, Table 2). The reason for substituting the longer T7-TN3 primer in this PCR step was that background amplification which occurred when using the shorter TN3 primer was suppressed effectively, particularly when the T7-TN3 primer was used at a low concentration (0.1 μ M) (Frohman et al., (1998) *PNAS USA*, 85, 8998-9002). The second PCR step involved the TN3 primer (SEQ ID No. 1, Table 2) and a second degenerate primer (Table 4).

TABLE 4

Combinations of Degenerate Primers for First and Second PCR Resulting in Specific Amplification of 3'-Fragments of nFP cDNA

| Species | First Degenerate Primer | Second Degenerate Primer |
|---------------------|-------------------------|--|
| Anemonia majano | NGH (SEQ ID No. 4) | GNGb (SEQ ID No. 10) |
| Clavularia sp. | NGH (SEQ ID No. 4) | GEGa (SEQ ID No. 6) |
| Zoanthus sp. | NGH (SEQ ID No. 4) | GEGa (SEQ ID No. 6) |
| Discosoma sp. "red" | NGH (SEQ ID No. 4) | GEGa (SEQ ID No. 6), NFP (SEQ ID No. 13) or PVMb (SEQ ID No. 16) |
| Discosoma striata | NGH (SEQ ID No. 4) | NFP (SEQ ID No. 13) |
| Anemonia sulcata | NGH (SEQ ID No. 4) | GEGa (SEQ ID No. 6) or NFP (SEQ ID No. 13) |

5

The first PCR reaction was performed as follows: 1 μ l of 20-fold
 10 dilution of the amplified cDNA sample was added into the reaction
 mixture containing 1X Advantage KlenTaq Polymerase Mix with
 provided buffer (CLONTECH), 200 μ M dNTPs, 0.3 μ M of first degenerate

primer (Table 4) and 0.1 μ M of T7-TN3 (SEQ ID No. 17) primer in a total volume of 20 μ l. The cycling profile was (Hybaid OmniGene Thermocycler, tube control mode): 1 cycle for 95°C, 10 sec.; 55°C, 1 min.; 72°C, 40 sec; 24 cycles for 95°C, 10 sec.; 62°C, 30 sec.; 72°C, 40 sec. The reaction was then diluted 20-fold in water and 1 μ l of this dilution was added to a second PCR reaction, which contained 1X Advantage KlenTaq Polymerase Mix with the buffer provided by the manufacturer (CLONTECH), 200 μ M dNTPs, 0.3 μ M of the second degenerate primer (Table 4) and 0.1 μ M of TN3 primer. The cycling profile was (Hybaid OmniGene Thermocycler, tube control mode): 1 cycle for 95°C, 10 sec.; 55°C (for GEG/GNG or PVM) or 52°C (for NFP), 1 min.; 72°C, 40 sec; 13 cycles for 95°C, 10sec.; 62°C (for GEG/GNG or PVM) or 58°C (for NFP), 30 sec.; 72°C, 40 sec. The product of PCR was cloned into PCR-Script vector (Stratagene) according to the manufacturer's protocol.

Different combinations of degenerate primers were tried in the first and second PCR reactions on the DNA from each species until a combination of primers was found that resulted in specific amplification--meaning that a pronounced band of expected size (about 650-800 bp for NGH and GEG/GNG and 350-500 bp for NFP and PVM--sometimes accompanied by a few minor bands) was detected on agarose gel after two PCR reactions. The primer combinations of choice for different species of the Class Anthozoa are listed in Table 4. Some other primer combinations also resulted in amplification of fragments of correct size, but the sequence of these fragments showed no homology to the other fluorescent proteins identified or to *Aequorea victoria* GFP.

EXAMPLE 5Obtaining Full-Length cDNA Copies

Upon sequencing the obtained 3'-fragments of novel
5 fluorescent protein cDNAs, two nested 5'-directed primers were
synthesized for cDNA (Table 5), and the 5' ends of the cDNAs were
then amplified using two consecutive PCRs. In the next PCR reaction,
the novel approach of "step-out PCR" was used to suppress background
amplification. The step-out reaction mixture contained 1x Advantage
10 KlenTaq Polymerase Mix using buffer provided by the manufacturer
(CLONTECH), 200 μ M dNTPs, 0.2 μ M of the first gene-specific primer
(see Table 5), 0.02 μ M of the T7-TS primer (SEQ ID No. 18), 0.1 μ M of
T7 primer (SEQ ID No. 19) and 1 μ l of the 20-fold dilution of the
amplified cDNA sample in a total volume of 20 μ l. The cycling profile
15 was (Hybaid OmniGene Thermocycler, tube control mode): 23-27
cycles for 95°C, 10 sec.; 60°C, 30 sec.; 72°C, 40 sec. The product of
amplification was diluted 50-fold in water and one μ l of this dilution
was added to the second (nested) PCR. The reaction contained 1X
Advantage KlenTaq Polymerase Mix with provided buffer (CLONTECH),
20 200 μ M dNTPs, 0.2 μ M of the second gene-specific primer and 0.1 μ M
of TS primer (SEQ ID No. 2) in a total volume of 20 μ l. The cycling
profile was (Hybaid OmniGene Thermocycler, tube control mode): 12
cycles for 95°C, 10 sec.; 60°C, 30 sec.; 72°C, 40 sec. The product of
amplification was then cloned into pAtlas vector (CLONTECH) according
25 to the manufacturer's protocol.

TABLE 5Gene-Specific Primers Used for 5'-RACE

| Species | First Primer | Second (Nested) Primer |
|-------------------------------|---|--|
| Anemonia majano | 5'-GAAATAGTCAGGCATACTGGT (SEQ ID No. 20) | 5'-GTCAGGCATAC TGGTAGGAT (SEQ ID No. 21) |
| Clavularia sp. | 5'-CTTGAAATAGTCTGCTATATC — (SEQ ID No. 22) | 5'-TCTGCTATATC GTCTGGGT (SEQ ID No. 23) |
| Zoanthus sp. | 5'- GTTCTTGAAATAGTCTACTATGT (SEQ ID No. 24) | 5'-GTCTACTATGTCTT GAGGAT (SEQ ID No. 25) |
| Discosoma sp. "red" | 5'-CAAGCAAATGGCAAAGGTC (SEQ ID No. 26) | 5'-CGGTATTGTGGCC TTCGTA (SEQ ID No. 27) |
| Discosoma striata | 5'-TTGTCTTCTTCTGCACAAC (SEQ ID No. 28) | 5'-CTGCACAACGG GTCCAT (SEQ ID No. 29) |
| Anemonia sulcata | 5'-CCTCTATCTTCATTTCCTGC (SEQ ID No. 30) | 5'-TATCTTCATTTCCT GCGTAC (SEQ ID No. 31) |
| Discosoma sp. "magenta" | 5'-TTCAGCACCCCATCACGAG (SEQ ID No. 32) | 5'-ACGCTCAGAGCTG GGTTCC (SEQ ID No. 33) |
| Discosoma sp. "green" | 5'-CCCTCAGCAATCCATCACGTTC (SEQ ID No. 34) | 5'-ATTATCTCAGTGGA TGGTTC (SEQ ID No. 35) |

EXAMPLE 6

Expression of nFP in *E. coli*

5 To prepare a DNA construct for novel fluorescent protein expression, two primers were synthesized for each cDNA: a 5'-directed "downstream" primer with the annealing site located in the 3'-UTR of the cDNA and a 3'-directed "upstream" primer corresponding to the site of translation start site (not including the first ATG codon) (Table 6). Primers with SEQ ID Nos. 38 and 40 or SEQ ID Nos. 39 and 40 were the primers used to prepare the c484 DNA. Both primers had 5'-heels coding for a site for a restriction endonuclease; in addition, the upstream primer was designed so as to allow the cloning of the PCR product into the pQE30 vector (Qiagen) in such a way that resulted in the fusion of reading frames of the vector-encoded 6xHis-tag and nFP. The PCR was performed as follows: 1 µl of the 20-fold dilution of the amplified cDNA sample was added to a mixture containing 1x Advantage KlenTaq Polymerase Mix with buffer provided by the manufacturer (CLONTECH), 200 µM dNTPs, 0.2 µM of upstream primer and 0.2 µM of downstream primer, in a final total volume of 20 µl. The cycling profile was (Hybaid OmniGene Thermocycler, tube control mode): 23-27 cycles for 95°C, 10 sec.; 60°C, 30 sec.; 72°C, 40 sec. The product of this amplification step was purified by phenol-chlorophorm extraction and ethanol precipitation and then cloned into pQE30 vector using restriction endonucleases corresponding to the primers' sequence according to standard protocols.

All plasmids were amplified in XL-1 blue *E. coli* and purified by plasmid DNA miniprep kits (CLONTECH). The recombinant clones were selected by colony color, and grown in 3 ml of LB medium

(supplemented with 100 µg/ml of ampicillin) at 37°C overnight. 100 µl of the overnight culture was transferred into 200 ml of fresh LB medium containing 100 µg/ml of ampicillin and grown at 37°C, 200 rpm up to OD₆₀₀ 0.6-0.7. 1 mM IPTG was then added to the culture and
5 incubation was allowed to proceed at 37°C for another 16 hours. The cells were harvested and recombinant protein, which incorporated 6x His tags on the N-terminus, was purified using TALON™ metal-affinity resin according to the manufacturer's protocol (CLONTECH).

TABLE 6

Primers Used to Obtain Full Coding Region of nEPs for Cloning into Expression Construct

| Species | Upstream Primer | Downstream Primer |
|-------------------------|--|---|
| Anemonia majano | 5' -acatggatccgctctttcaaaca agtttacc (SEQ ID No. 36) BamHI | 5'-tagtactcgagcttattcgta tttcagtgaatc (SEQ ID No. 37) XhoI |
| Clavularia sp. | L: 5'-acatggatccaacattttttga gaaacg (SEQ ID No. 38) BamHI S: 5'-acatggatccaaagctctaacc accatg (SEQ ID No. 39) BamHI | 5'-tagtactcgagcaacacaa accctcagacaa (SEQ ID No. 40) XhoI |
| Zoanthus sp. | 5'- acatggatccgctcagtcacaag cacggt (SEQ ID No. 41) BamHI | 5'-tagtactcgagggttggaactacat tcttata (SEQ ID No. 42) XhoI |
| Discosoma sp. "red" | 5'- acatggatccaggtcttccaagaat gttacc (SEQ ID No. 43) BamHI | 5'-tagtactcgaggagccaagttc agcctta (SEQ ID No. 44) XhoI |
| Discosoma striata | 5'- acatggatccagttggtccaagagtgtg (SEQ ID No. 45) BamHI | 5'-tagcgagctctatcatgcctc gtcacct (SEQ ID No. 46) SacI |
| Anemonia sulcata | 5'- acatggatccgcttccttttaagaagact (SEQ ID No. 47) BamHI | 5'-tagtactcgagtccttgggagc ggcttg (SEQ ID No. 48) XhoI |
| Discosoma sp. "magenta" | 5'- acatggatccagttgtccaagaatgtgat (SEQ ID No. 49) BamHI | 5'-tagtactcgaggccattacg ctaacc (SEQ ID No. 50) XhoI |
| Discosoma sp. "green" | 5'-acatggatccagtgcaactaaagaagaaatg (SEQ ID No. 51) | 5'-tagtactcgagattcggtttaat gccttg (SEQ ID No. 52) |

EXAMPLE 7Novel Fluorescent Proteins and cDNAs Encoding the Proteins

One of the full-length cDNAs encoding fluorescent proteins
 5 found is described herein (cFP484). The nucleic acid sequence and
 deduced amino acid sequence are SEQ ID Nos. 55 and 56, respectively.
 The spectral properties of cFP484 is listed in Table 7, and the emission
 and excitation spectra for the cFP484 is shown in Figure 2.

10 **TABLE 7**

Spectral Properties of the Isolated cFP484 nFP

| | | | |
|---------------------------|-------------------|---------------------------------|--------|
| Species: | Clavularia sp. | Max. Extinction Coefficient: | 35,300 |
| 15 nFP Name: | cFP484 | Quantum Yield | 0.48 |
| Absorbance Max. (nm): | 456 | Relative Brightness:* | 0.77 |
| 20 Emission Max. (nm): | 484 | | |

*relative brightness is extinction coefficient multiplied by quantum
 yield divided by the same value for *A. victoria* GFP.

25 **EXAMPLE 8**

Generation of Deletion Mutants of cFP484

Two deletion mutants were generated by two separate PCR
 reactions: $\Delta 19$ cFP484 lacks the N-terminal first 19 amino acids of
 30 cFP484, and $\Delta 38$ cFP484 lacks the N-terminal first 38 amino acids of
 cFP484. Mammalian expression vectors containing the DNA encoding

the fluorescent protein $\Delta 19$ cFP484 or $\Delta 38$ cFP484 are generated, which are named as p $\Delta 19$ NFP2-N1 and p $\Delta 38$ NFP2-N1, respectively.

5

EXAMPLE 9

Transient Expression of Deletion Mutants of cFP484 in Mammalian Cells

HeLa cells were transiently transfected with mammalian expression vector p $\Delta 19$ NFP2-N1 which contains the DNA encoding the fluorescent protein $\Delta 19$ cFP484. After transfection, cells were incubated for 48 hours at 37°C then fixed in 3.7% formaldehyde. Cells were mounted in mounting medium and observed by fluorescence microscopy. Digital images were taken with MetaMorph software (Universal Imaging Corp.) using a monochrome cooled CCD camera (Roper Scientific). The filter set XF 114 (Omega Optical) was used to visualize fluorescence emitted by $\Delta 19$ cFP484. The image was pseudocolored (Figure 3A). Shown in Figure 3B is phase contrast image taken from the same field of view as that in Figure 3A and overlaid with image in Figure 3A.

$\Delta 38$ cFP484 is also fluorescent when expressed in HeLa cells (Data not shown).

Any patents or publications mentioned in this specification are indicative of the levels of those skilled in the art to which the invention pertains. These patents and publications are incorporated by reference to the same extent as if each individual publication was specifically and individually indicated to be incorporated by reference.

One skilled in the art will appreciate readily that the present invention is adapted to carry out the objects and obtain the ends and

advantages mentioned, as well as those objects and ends inherent therein. The present examples, along with the methods, procedures, treatments, molecules, and specific compounds described herein, are presently representative of preferred embodiments, are exemplary, and
5 are not intended as limitations on the scope of the invention. Changes to the methods and compounds, and other uses, will occur to those skilled in the art and are encompassed within the spirit of the invention as defined by the scope of the claims.

WHAT IS CLAIMED IS:

1. A DNA sequence encoding a fluorescent protein selected from the group consisting of:

5 (a) an isolated DNA which encodes a fluorescent protein, wherein said DNA is from an organism from a Class Anthozoa and wherein said organism does not exhibit bioluminescence;

(b) an isolated DNA which hybridizes to isolated DNA of (a) above and which encodes a fluorescent protein; and

10 (c) an isolated DNA differing from the isolated DNAs of (a) and (b) above in codon sequence due to degeneracy of the genetic code and which encodes a fluorescent protein.

2. The DNA sequence of claim 1, wherein said organism
15 is from Sub-class Alcyonaria.

3. The DNA sequence of claim 2, wherein said organism is from Order Stolonifera.

20 4. The DNA sequence of claim 3, wherein said organism is from Family Clavulariidae.

5. The DNA sequence of claim 4, wherein said organism is from Genus Clavularia.

25

6. A DNA sequence encoding a fluorescent protein selected from the group consisting of:

(a) an isolated DNA which encodes a fluorescent protein, wherein said DNA has a sequence shown in SEQ ID No. 55;

(b) an isolated DNA which hybridizes to isolated DNA of (a) above and which encodes a fluorescent protein; and

(c) an isolated DNA differing from the isolated DNAs of (a) and (b) above in codon sequence due to degeneracy of the genetic
5 code, and which encodes a fluorescent protein.

7. The DNA of claim 6, wherein said DNA encodes a fluorescent protein having an amino acid sequence shown in SEQ ID No.
56.
10

8. A vector capable of expressing the DNA of claim 1 in a recombinant cell, said vector comprising said DNA of claim 1 and regulatory elements necessary for expression of the DNA in the cell.

15 9. The vector of claim 8, wherein said DNA encodes a fluorescent protein having the amino acid sequence shown in SEQ ID No. 56.

10. A host cell transfected with the vector of claim 8,
20 wherein said cell is capable of expressing a fluorescent protein.

11. The host cell of claim 10, wherein said cell is selected from the group consisting of bacterial cells, mammalian cells, plant cell, yeast and insect cells.
25

12. The host cell of claim 11, wherein said bacterial cell is an *E. coli* cell.

13. An isolated and purified fluorescent protein coded for by DNA selected from the group consisting of:

(a) an isolated DNA which encodes a fluorescent protein from an organism from Class Anthozoa, wherein said organism does not exhibit bioluminescence;

(b) an isolated DNA which hybridizes to isolated DNA of (a) above and which encodes a fluorescent protein; and

(c) an isolated DNA differing from the isolated DNAs of (a) and (b) above in codon sequence due to degeneracy of the genetic code and which encodes a fluorescent protein.

14. The isolated and purified fluorescent protein of claim 13, wherein said organism is from Sub-class Alcyonaria.

15. The isolated and purified fluorescent protein of claim 14, wherein said organism is from Order Stolonifera.

16. The isolated and purified fluorescent protein of claim 15, wherein said organism is from Family Clavulariidae.

17. The isolated and purified fluorescent protein of claim 16, wherein said organism is from Genus Clavularia.

18. An isolated and purified fluorescent protein coded for by DNA selected from the group consisting of:

(a) isolated DNA which encodes a fluorescent protein having an amino acid sequence shown in SEQ ID No. 56;

(b) isolated DNA which hybridizes to isolated DNA of (a) above and which encodes a fluorescent protein; and

(c) isolated DNA differing from said isolated DNAs of (a) and (b) above in codon sequence due to degeneracy of the genetic code and which encodes a fluorescent protein.

5 19. The isolated and purified fluorescent protein of claim 18, wherein said protein is cFP484.

 20. An amino acid sequence which can be used as a basis for designing an oligonucleotide probe for identification of a DNA
10 encoding a fluorescent protein by means of hybridization, wherein said sequence is selected from the group consisting of SEQ ID Nos. 3, 5, 8, 11, 12, 14.

 21. The amino acid sequence of claim 20, wherein said
15 oligonucleotide has a nucleotide sequence selected from the group consisting of SEQ ID Nos. 4, 6, 7, 9, 10, 13, 15, 16

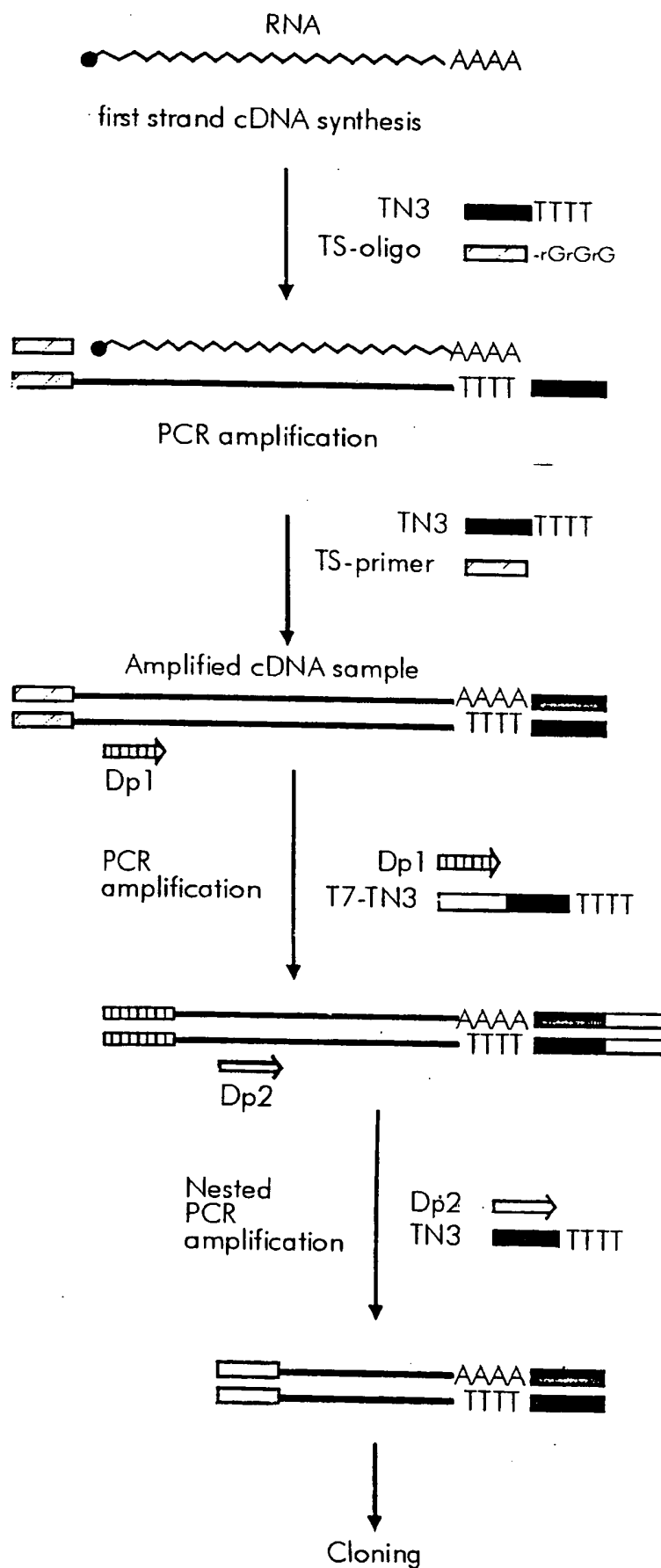


FIG. 1

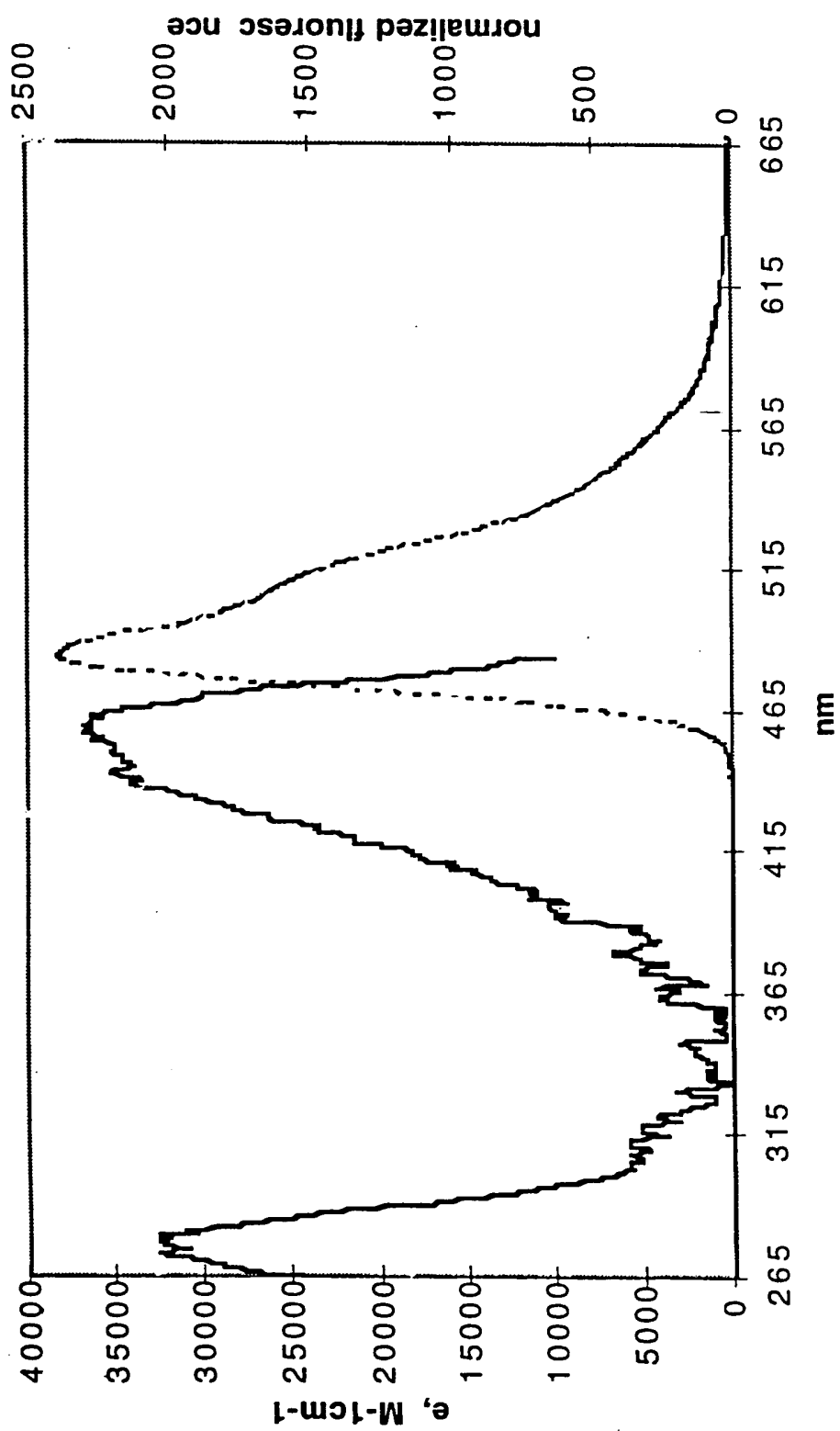


FIG. 2

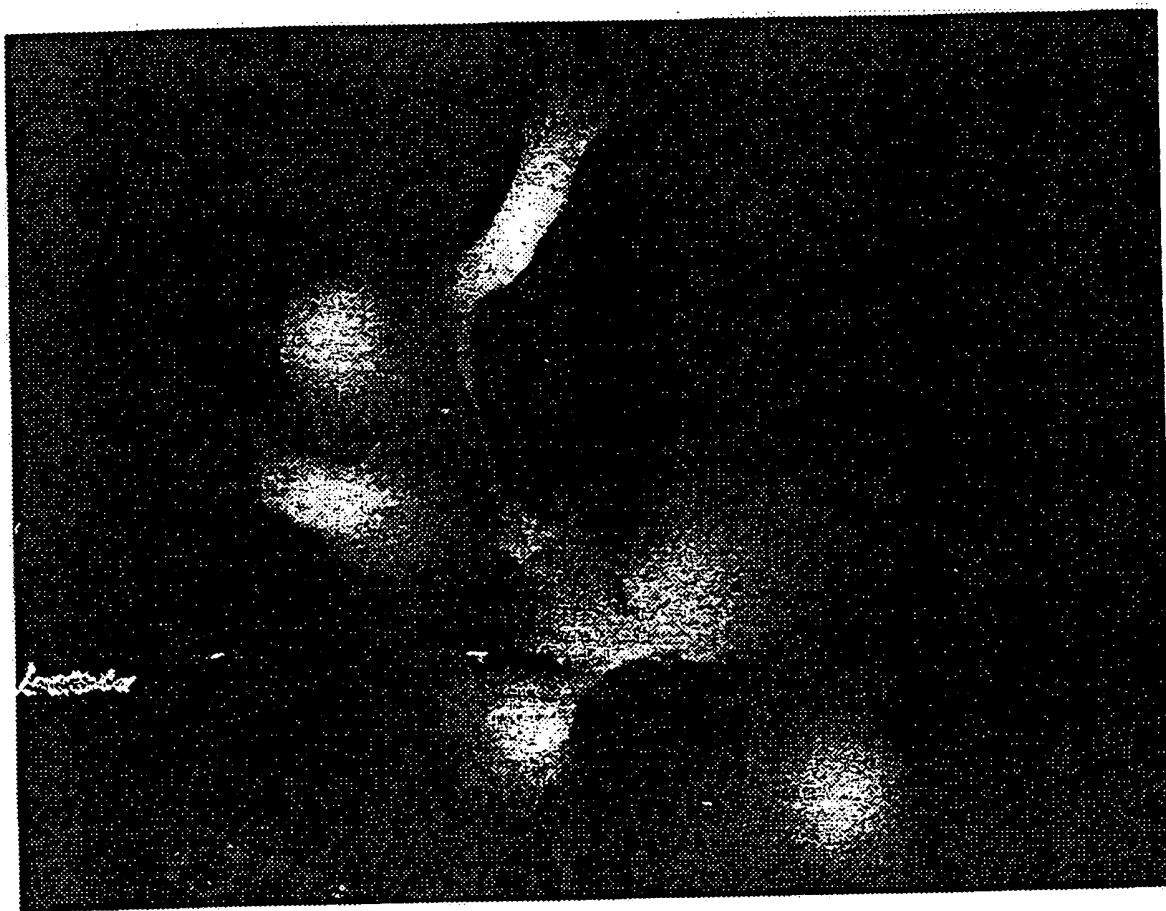


FIG. 3A

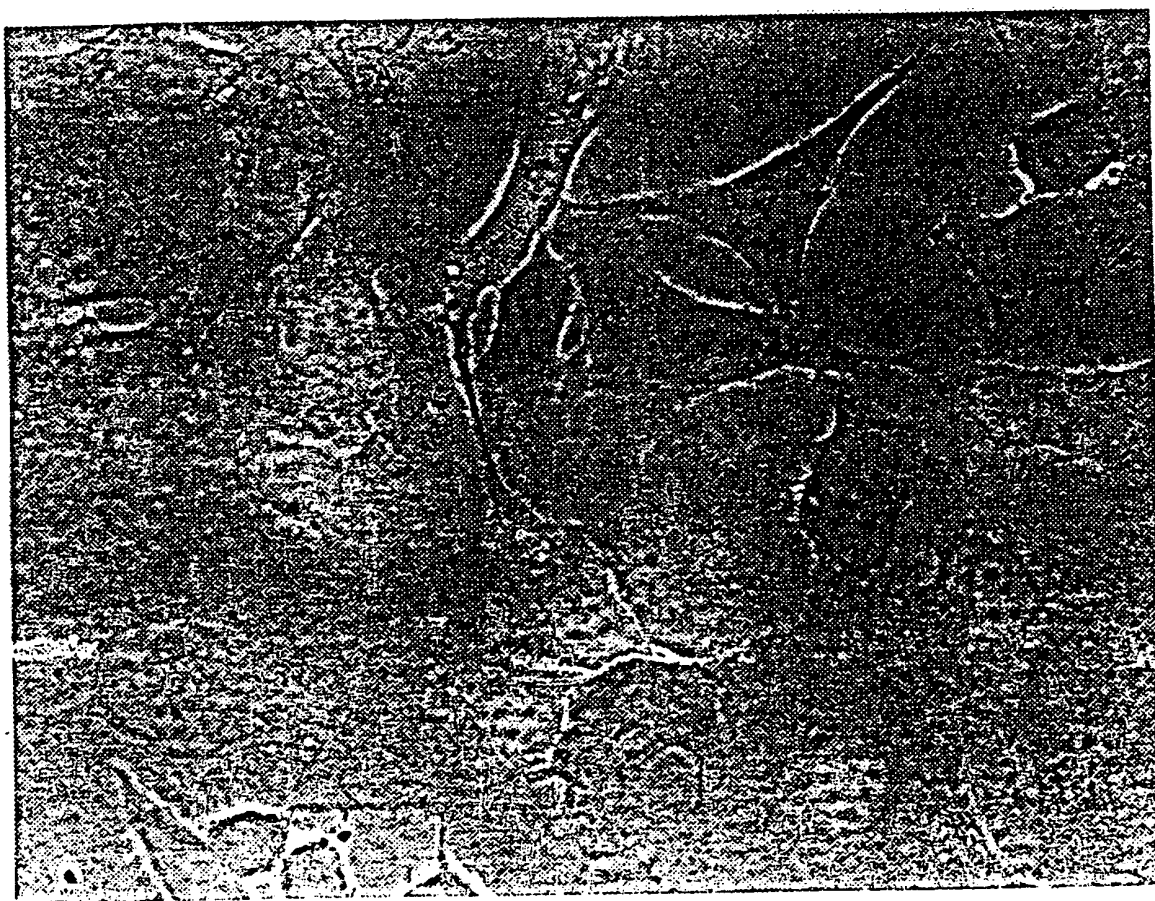


FIG. 3B

SEQUENCE LISTING

<110> Lukyanov, Sergey A.
Labas, Yulii A.
Matz, Mikhail V.
Fradkov, Arcady F.
Angres, Brigitte
Green, Gisele

<120> Fluorescent Proteins from Non-Bioluminescent
Species of Class Anthozoa, Genes Encoding Such
Proteins and Uses Thereof

<130> D6196D2PCT

<140> 09/418,917

<141> 1999-10-15

<150> 09/210,330

<151> 1998-12-11

<160> 56

<210> 1

<211> 25

<212> DNA

<213> artificial sequence

<220>

<221> primer_bind

<223> primer TN3 used in cDNA synthesis and RACE

<400> 1

cgcagtcgac cgttttttttt ttttt 25

<210> 2

<211> 23

<212> DNA

<213> artificial sequence

<220>

<221> primer_bind

<223> primer TS used in cDNA synthesis and RACE

<400> 2

aagcagtggc atcaacgcag agt 23

<210> 3

<211> 6
 <212> PRT
 <213> *Aequorea victoria*
 <220>
 <222> 21
 <223> amino acid sequence of a key stretch on which
 primer NGH is based; Xaa at position 21 represents
 unknown
 <400> 3

Gly Xaa Val Asn Gly His

5

<210> 4
 <211> 20
 <212> DNA
 <213> artificial sequence
 <220>
 <221> primer_bind
 <222> 12
 <223> primer NGH used for isolation of fluorescent
 protein; n at position 12 represents any of the
 four bases
 <400> 4

gayggctgcg tnaayggdca

20

<210> 5
 <211> 5
 <212> PRT
 <213> *Aequorea victoria*
 <220>
 <222> 31...35
 <223> amino acid sequence of a key stretch on which
 primers GEGa and GEGb are based
 <400> 5

Gly Glu Gly Glu Gly

5

<210> 6
<211> 20
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> primer GEGa used for isolation of fluorescent
protein
<400> 6

gttacaggtg arggmgargg 20

<210> 7
<211> 20
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> primer GEGb used for isolation of fluorescent
protein
<400> 7

gttacaggtg arggkgargg 20

<210> 8
<211> 5
<212> PRT
<213> *Aequorea victoria*
<220>
<222> 31...35
<223> amino acid sequence of a key stretch on which
primers GNGa and GNGb are based
<400> 8

Gly Glu Gly Asn Gly
5

<210> 9
<211> 20
<212> DNA

<213> artificial sequence
<220>
<221> primer_bind
<223> primer GNGa used for isolation of fluorescent
protein
<400> 9

gttacaggtg arggmaaygg 20

<210> 10
<211> 20
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> primer GNGb used for isolation of fluorescent
protein
<400> 10

gttacaggtg arggkaaygg 20

<210> 11
<211> 5
<212> PRT
<213> *Aequorea victoria*
<220>
<222> 127..131
<223> amino acid sequence of a key stretch on which
primer NFP is based
<400> 11

Gly Met Asn Phe Pro

5

<210> 12
<211> 5
<212> PRT
<213> *Aequorea victoria*
<220>
<222> 127..131

<223> amino acid sequence of a key stretch on which
primer NFP is based

<400> 12

Gly Val Asn Phe Pro

5

<210> 13

<211> 20

<212> DNA

<213> artificial sequence

<220>

<221> primer_bind

<223> primer NFP used for isolation of fluorescent
protein

<400> 13

ttccayggtr tgaayttycc

20

<210> 14

<211> 4

<212> PRT

<213> *Aequorea victoria*

<220>

<222> 134...137

<223> amino acid sequence of a key stretch on which
primers PVMa and PVMb are based

<400> 14

Gly Pro Val Met

<210> 15

<211> 21

<212> DNA

<213> artificial sequence

<220>

<221> primer_bind

<222> 15

<223> primer PVMa used for isolation of fluorescent protein; n at position 15 represents any of the four bases

<400> 15

cctgccrayg gtcnngtmat g

21

<210> 16

<211> 21

<212> DNA

<213> artificial sequence

<220>

<221> primer_bind

<222> 15

<223> primer PVMb used for isolation of fluorescent protein; n at position 15 represents any of the four bases

<400> 16

cctgccrayg gtcnngtkat g

21

<210> 17

<211> 47

<212> DNA

<213> artificial sequence

<220>

<221> primer_bind

<223> primer T7-TN3 used in cDNA synthesis and RACE

<400> 17

gtaatacgac tcactatagg gccgcagtcg accgtttttt ttttttt

47

<210> 18

<211> 45

<212> DNA

<213> artificial sequence

<220>

<221> primer_bind

<223> primer T7-TS used in cDNA synthesis and RACE

<400> 18

gtaatacgac tcactatagg gcaagcagtg gtatcaacgc agagt

45

<210> 19
<211> 22
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> primer T7 used in cDNA synthesis and RACE
<400> 19

gtaatacgac tcactatagg gc

22

<210> 20
<211> 21
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> gene-specific primer used for 5'-RACE for *Anemonia majano*
<400> 20

gaaatagtca ggcatactgg t

21

<210> 21
<211> 20
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> gene-specific primer used for 5'-RACE for *Anemonia majano*
<400> 21

gtcaggcata ctggtaggat

20

<210> 22

<211> 21
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> gene-specific primer used for 5'-RACE for
Clavularia sp.
<400> 22

cttgaaatag tctgctatat c 21

<210> 23
<211> 19
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> gene-specific primer used for 5'-RACE for
Clavularia sp.
<400> 23

tctgctatat cgtctgggt 19

<210> 24
<211> 23
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> gene-specific primer used for 5'-RACE for *Zoanthus*
sp.
<400> 24

gttcttgaaa tagtctacta tgt 23

<210> 25
<211> 20
<212> DNA

<213> artificial sequence
<220>
<221> primer_bind
<223> gene-specific primer used for 5'-RACE for *Zoanthus*
sp.
<400> 25

gtctactatg tcttgaggat 20

<210> 26
<211> 19
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> gene-specific primer used for 5'-RACE for
Discosoma sp. "red"
<400> 26

caagcaaagtc gcaaaggctc 19

<210> 27
<211> 19
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> gene-specific primer used for 5'-RACE for
Discosoma sp. "red"
<400> 27

cggtattgtg gccttcgta 19

<210> 28
<211> 19
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind

<223> gene-specific primer used for 5'-RACE for
Discosoma striata

<400> 28

ttgtcttctt ctgcacaac 19

<210> 29

<211> 17

<212> DNA

<213> artificial sequence

<220>

<221> primer_bind

<223> gene-specific primer used for 5'-RACE for
Discosoma striata

<400> 29

ctgcacaacg ggtccat 17

<210> 30

<211> 20

<212> DNA

<213> artificial sequence

<220>

<221> primer_bind

<223> gene-specific primer used for 5'-RACE for *Anemonia*
sulcata

<400> 30

cctctatctt catttctg 20

<210> 31

<211> 20

<212> DNA

<213> artificial sequence

<220>

<221> primer_bind

<223> gene-specific primer used for 5'-RACE for *Anemonia*
sulcata

<400> 31

tatcttcatt tcctgcgtac

20

<210> 32

<211> 19

<212> DNA

<213> artificial sequence

<220>

<221> primer_bind

<223> gene-specific primer used for 5'-RACE for
Discosoma sp. "magenta"

<400> 32

ttcagcaccc catcacgag

19

<210> 33

<211> 19

<212> DNA

<213> artificial sequence

<220>

<221> primer_bind

<223> gene-specific primer used for 5'-RACE for
Discosoma sp. "magenta"

<400> 33

acgctcagag ctgggttcc

19

<210> 34

<211> 22

<212> DNA

<213> artificial sequence

<220>

<221> primer_bind

<223> gene-specific primer used for 5'-RACE for
Discosoma sp. "green"

<400> 34

ccctcagcaa tccatcacgt tc

22

<210> 35

<211> 20
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> gene-specific primer used for 5'-RACE for
Discosoma sp. "green"
<400> 35

attatctcag tggatggttc 20

<210> 36
<211> 31
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> upstream primer used to obtain full coding region
of nFPs from *Anemonia majano*
<400> 36

acatggatcc gctctttcaa acaagtttat c 31

<210> 37
<211> 34
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> downstream primer used to obtain full coding
region of nFPs from *Anemonia majano*
<400> 37

tagtactcga gcttattcgt atttcagtga aatc 34

<210> 38
<211> 29
<212> DNA
<213> artificial sequence
<220>

<221> primer_bind
<223> upstream primer used to obtain full coding region
of nFPs from *Clavularia sp.*
<400> 38

acatggatcc aacatttttt tgagaaacg 29

<210> 39
<211> 28
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> upstream primer used to obtain full coding region
of nFPs from *Clavularia sp.*
<400> 39

acatggatcc aaagctctaa ccaccatg 28

<210> 40
<211> 31
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> downstream primer used to obtain full coding
region of nFPs from *Clavularia sp.*
<400> 40

tagtactcga gcaacacaaa ccctcagaca a 31

<210> 41
<211> 28
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> upstream primer used to obtain full coding region
of nFPs from *Zoanthus sp.*

<400> 41

acatggatcc gctcagtcaa agcacggt 28

<210> 42

<211> 32

<212> DNA

<213> artificial sequence

<220>

<221> primer_bind

<223> downstream primer used to obtain full coding
region of nFPs from *Zoanthus* sp.

<400> 42

tagtactcga gggtggaact acattcttat ca 32

<210> 43

<211> 31

<212> DNA

<213> artificial sequence

<220>

<221> primer_bind

<223> upstream primer used to obtain full coding region
of nFPs from *Discosoma* sp. "red"

<400> 43

acatggatcc aggtcttcca agaatgttat c 31

<210> 44

<211> 29

<212> DNA

<213> artificial sequence

<220>

<221> primer_bind

<223> downstream primer used to obtain full coding
region of nFPs from *Discosoma* sp. "red"

<400> 44

tagtactcga ggagccaagt tcagcctta 29

<210> 45
<211> 28
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> upstream primer used to obtain full coding region
of nFPs from *Discosoma striata*
<400> 45

acatggatcc agttggtcca agagtgtg 28

<210> 46
<211> 28
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> downstream primer used to obtain full coding
region of nFPs from *Discosoma striata*
<400> 46

tagcgagctc tatcatgcct cgtcacct 28

<210> 47
<211> 31
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> upstream primer used to obtain full coding region
of nFPs from *Anemonia sulcata*
<400> 47

acatggatcc gcttcctttt taaagaagac t 31

<210> 48
<211> 28
<212> DNA
<213> artificial sequence

<220>

<221> primer_bind

<223> downstream primer used to obtain full coding
region of nFPs from *Anemonia sulcata*

<400> 48

tagtactcga gtccttgga gcggttg

28

<210> 49

<211> 30

<212> DNA

<213> artificial sequence

<220>

<221> primer_bind

<223> upstream primer used to obtain full coding region
of nFPs from *Discosoma sp. "magenta"*

<400> 49

acatggatcc agttgtcca agaattgat

30

<210> 50

<211> 26

<212> DNA

<213> artificial sequence

<220>

<221> primer_bind

<223> downstream primer used to obtain full coding
region of nFPs from *Discosoma sp. "magenta"*

<400> 50

tagtactcga ggccattacg ctaatc

26

<210> 51

<211> 31

<212> DNA

<213> artificial sequence

<220>

<221> primer_bind

<223> upstream primer used to obtain full coding region
of nFPs from *Discosoma sp. "green"*

<400> 51
 acatggatcc agtgcactta aagaagaaat g 31

<210> 52
 <211> 29
 <212> DNA
 <213> artificial sequence
 <220>
 <221> primer_bind
 <223> downstream primer used to obtain full coding
 region of nFPs from *Discosoma sp.* "green"
 <400> 52
 tagtactcga gattcggttt aatgccttg 29

<210> 53
 <211> 33
 <212> DNA
 <213> artificial sequence
 <220>
 <221> primer_bind
 <223> TS-oligo used in cDNA synthesis and RACE
 <400> 53
 aagcagtggg atcaacgcag agtacgcrgr grg 33

<210> 54
 <211> 238
 <212> PRT
 <213> *Aequorea victoria*
 <220>
 <223> amino acid sequence of GFP
 <400> 54

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Ser | Lys | Gly | Glu | Glu | Leu | Phe | Thr | Gly | Val | Val | Pro | Ile | Leu |
| | | | | 5 | | | | | 10 | | | | | 15 |
| Val | Glu | Leu | Asp | Gly | Asp | Val | Asn | Gly | His | Lys | Phe | Ser | Val | Ser |
| | | | | 20 | | | | | 25 | | | | | 30 |
| Gly | Glu | Gly | Glu | Gly | Asp | Ala | Thr | Tyr | Gly | Lys | Leu | Thr | Leu | Lys |
| | | | | 35 | | | | | 40 | | | | | 45 |

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Phe | Ile | Cys | Thr | Thr | Gly | Lys | Leu | Pro | Val | Pro | Trp | Pro | Thr | Leu |
| | | | | 50 | | | | | 55 | | | | | 60 |
| Val | Thr | Thr | Phe | Ser | Tyr | Gly | Val | Gln | Cys | Phe | Ser | Arg | Tyr | Pro |
| | | | | 65 | | | | | 70 | | | | | 75 |
| Asp | His | Met | Lys | Gln | His | Asp | Phe | Phe | Lys | Ser | Ala | Met | Pro | Glu |
| | | | | 80 | | | | | 85 | | | | | 90 |
| Gly | Tyr | Val | Gln | Glu | Arg | Thr | Ile | Phe | Phe | Lys | Asp | Asp | Gly | Asn |
| | | | | 95 | | | | | 100 | | | | | 105 |
| Tyr | Lys | Thr | Arg | Ala | Glu | Val | Lys | Phe | Glu | Gly | Asp | Thr | Leu | Val |
| | | | | 110 | | | | | 115 | | | | | 120 |
| Asn | Arg | Ile | Glu | Leu | Lys | Gly | Ile | Asp | Phe | Lys | Glu | Asp | Gly | Asn |
| | | | | 125 | | | | | 130 | | | | | 135 |
| Ile | Leu | Gly | His | Lys | Leu | Glu | Tyr | Asn | Tyr | Asn | Ser | His | Asn | Val |
| | | | | 140 | | | | | 145 | | | | | 150 |
| Tyr | Ile | Met | Ala | Asp | Lys | Gln | Lys | Asn | Gly | Ile | Lys | Val | Asn | Phe |
| | | | | 155 | | | | | 160 | | | | | 165 |
| Lys | Ile | Arg | His | Asn | Ile | Glu | Asp | Gly | Ser | Val | Gln | Leu | Ala | Asp |
| | | | | 170 | | | | | 175 | | | | | 180 |
| His | Tyr | Gln | Gln | Asn | Thr | Pro | Ile | Gly | Asp | Gly | Pro | Val | Leu | Leu |
| | | | | 185 | | | | | 190 | | | | | 195 |
| Pro | Asp | Asn | His | Tyr | Leu | Ser | Thr | Gln | Ser | Ala | Leu | Ser | Lys | Asp |
| | | | | 200 | | | | | 205 | | | | | 210 |
| Pro | Asn | Glu | Lys | Arg | Asp | His | Met | Val | Leu | Leu | Glu | Phe | Val | Thr |
| | | | | 215 | | | | | 220 | | | | | 225 |
| Ala | Ala | Gly | Ile | Thr | His | Gly | Met | Asp | Glu | Leu | Tyr | Lys | | |
| | | | | 230 | | | | | 235 | | | | | |

<210> 55

<211> 1116

<212> DNA

<213> *Clavularia sp.*

<220>

<221> CDS

<222> 8, 12, 13, 17, 65, 77, 102, 104

<223> cDNA sequence of cFP484; n represents unknown

<400> 55

tataggancca tnngggngat tgggggtccaa agcattgtaa ccaacgcaga 50
 taacccccag tgggtntcaaa cgcaganaac gcgggaacat tggaaaattg 100


```

antnttaagg aggcaaggaa tcgggagtaa agttgcgaga aactgaaaaa 150
atgaagtgtgta aatttgtgtt ctgcctgtcc ttcttggtcc tcgccatcac 200
aaacgcgaac atttttttga gaaacgaggc tgacttcgaa gagaagacat 250
ttagaataacc aaaagctcta accaccatgg gtgtgattaa accagacatg 300
aagattaagc tgaagatgga aggaaatgta aacgggcatg cttttgtgat 350
tgaaggagaa ggagaaggaa agccttacga tgggacacac actttaaac 400
tggaagtga ggaaggtgcg cctctgcctt tttcttacga tatcttgatca 450
aacgcgttcc agtacggaaa cagagcattg acaaaatacc cagacgatat 500
agcagactat ttcaagcagt cgtttcccgga gggatattcc tgggaaagaa 550
ccatgacttt tgaagacaaa ggcattgtca aagtgaaaag tgacataagc 600
atggaggaag actcctttat ctatgaaatt cgttttgatg ggatggactt 650
tcctcccaat ggtccggtta tgcagaaaaa aactttgaag tgggaaccat 700
ccactgagat tatgtacgtg cgtgatggag tgctggtcgg agatattagc 750
cattctctgt tgctggaggg aggtggccat taccgatgtg acttcaaaag 800
tatttacaaa gcaaaaaaag ttgtcaaatt gccagactat cactttgtgg 850
accatcgcat tgagatcttg aaccatgaca aggattacaa caaagtaacg 900
ctgtatgaga atgcagttgc tcgctattct ttgctgccaa gtcaggccta 950
gacaacaagg atactgaaaa catatttgtc tgagggtttg tgttgttttt 1000
taaaagacat cagctcagca ttcgttagtt gtaacaaaaa atagctttta 1050
tttttggtgg gattaaatca tagggatttg ttttagtaat cattttgctt 1100
aataaaaagt gccttg 1116

```

<210> 56

<211> 266

<212> PRT

<213> *Clavularia sp.*

<220>

<223> amino acid sequence of cFP484

<400> 56

```

Met Lys Cys Lys Phe Val Phe Cys Leu Ser Phe Leu Val Leu Ala
      5                      10                      15
Ile Thr Asn Ala Asn Ile Phe Leu Arg Asn Glu Ala Asp Phe Glu
      20                      25                      30
Glu Lys Thr Phe Arg Ile Pro Lys Ala Leu Thr Thr Met Gly Val
      35                      40                      45
Ile Lys Pro Asp Met Lys Ile Lys Leu Lys Met Glu Gly Asn Val
      50                      55                      60

```


| | | | |
|---|-----|-----|-----|
| Asn Gly His Ala Phe Val Ile Glu Gly Glu Gly Glu Gly Lys Pro | 65 | 70 | 75 |
| Tyr Asp Gly Thr His Thr Leu Asn Leu Glu Val Lys Glu Gly Ala | 80 | 85 | 90 |
| Pro Leu Pro Phe Ser Tyr Asp Ile Leu Ser Asn Ala Phe Gln Tyr | 95 | 100 | 105 |
| Gly Asn Arg Ala Leu Thr Lys Tyr Pro Asp Asp Ile Ala Asp Tyr | 110 | 115 | 120 |
| Phe Lys Gln Ser Phe Pro Glu Gly Tyr Ser Trp Glu Arg Thr Met | 125 | 130 | 135 |
| Thr Phe Glu Asp Lys Gly Ile Val Lys Val Lys Ser Asp Ile Ser | 140 | 145 | 150 |
| Met Glu Glu Asp Ser Phe Ile Tyr Glu Ile Arg Phe Asp Gly Met | 155 | 160 | 165 |
| Asp Phe Pro Pro Asn Gly Pro Val Met Gln Lys Lys Thr Leu Lys | 170 | 175 | 180 |
| Trp Glu Pro Ser Thr Glu Ile Met Tyr Val Arg Asp Gly Val Leu | 185 | 190 | 195 |
| Val Gly Asp Ile Ser His Ser Leu Leu Leu Glu Gly Gly Gly His | 200 | 205 | 210 |
| Tyr Arg Cys Asp Phe Lys Ser Ile Tyr Lys Ala Lys Lys Val Val | 215 | 220 | 225 |
| Lys Leu Pro Asp Tyr His Phe Val Asp His Arg Ile Glu Ile Leu | 230 | 235 | 240 |
| Asn His Asp Lys Asp Tyr Asn Lys Val Thr Leu Tyr Glu Asn Ala | 245 | 250 | 255 |
| Val Ala Arg Tyr Ser Leu Leu Pro Ser Gln Ala | 260 | 265 | |

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US99/29402

A. CLASSIFICATION OF SUBJECT MATTER:

US CL :

435/320.1, 252.3, 252.33, 325, 410, 254.11, 348, 369, 69.1; 530/350; 536/23.5

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

Dialog files 155, 5, 434, 34, 357, 28, 44, 35, 77 (Medline, Biosis, Scisearch, Derwent Biotech abs., Oceanic Abs., Aquatic Sci. & Fish Abs., Dissertation Abs. Online, Conference papers Index); STN-CAS files registry, CAPLUS; WEST, files USPT, Derwent WPI,
search terms: fluoresc?, bioluminesc?, protein? polypeptide?, anthozo?, zoanthar? , coral? Cnidar?, inverteb?, anemon?, alcyonar?, octocorall?, stolonif?, clavulari?, gene#, zoanthid?, mscsksvi/sqsp, vngh/sqep, gegeg/sqep, gegng/sqep, gmnfp/sqep, gvnfp/sqep, gpvn/sqep

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/29402

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|-----------|--|-----------------------|
| A | ANDERLUH et al. Cloning, sequencing, and expression of equinatoxin II. Biochemical and Biophysical Research Communications. 1996, Volume 220, No. 2, pages 437-442, entire document. | 1, 6, 8, 10-13, 18 |
| X | MACEK et al. Intrinsic tryptophan fluorescence of equinatoxin II, a pore-forming polypeptide from the sea anemone, <i>Actinia equina</i> | 13, 18 |
| --- | L, monitors its interaction with lipid membranes. European Journal of Biochemistry. 1995, Volume 234, pages 329-335, entire document. Cited as "I" document because it establishes fluorescence of equinatoxin II. | ----- |
| L | | 1, 6, 8, 10-12 |

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/29402

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C07K 14/435; C12N 1/00, 1/19, 5/10, 15/12, 15/63

US CL : Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/320.1, 252.3, 252.33, 325, 410, 254.11, 348, 369, 69.1; 530/350; 536/23.5

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|-----------|--|-----------------------|
| ***** | The sequence diskette submitted with the description was defective; thus the documents listed below were obtained solely by a word search. No SEQ ID NOs. could be searched. | ***** |
| X, P | MATZ et al. Fluorescent proteins from nonbioluminescent Anthozoa species. Nature Biotechnology. October 1999, Volume 17, No. 10, pages 969-673, entire document. | 1-21 |
| X, P | DE 197 18 640 A1 (WIEDENMANN) 22 July 1999, entire document. | 13, 18 |

☒ Further documents are listed in the continuation of Box C.
 ☐ See patent family annex.

| | |
|---|--|
| * Special categories of cited documents: | *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention |
| *A* document defining the general state of the art which is not considered to be of particular relevance | *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone |
| *E* earlier document published on or after the international filing date | *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | *G* document member of the same patent family |
| *O* document referring to an oral disclosure, use, exhibition or other means | |
| *P* document published prior to the international filing date but later than the priority date claimed | |

Date of the actual completion of the international search
06 MARCH 2000Date of mailing of the international search report
04 APR 2000Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

GABRIELE ELISABETH BUGAISKY

Telephone No. (703) 308-0196